Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Computer and Control Engineering (35^{th}cycle)

Study and implementation of new computational paradigms exploiting neuromorphic hardware architectures

By

Evelina Forno

******

Supervisors:
Prof. Gianvito Urgese
Prof. Enrico Macii

Doctoral Examination Committee:
Prof. Steve Furber, Referee, University of Manchester
Prof. Alejandro Linares Barranco, Referee, Universidad de Sevilla
Prof. Andrea Calimera, Politecnico di Torino
Prof. Egidio Falotico, Scuola Superiore Sant'Anna
Dr. James Knight, University of Sussex

Politecnico di Torino
2023

Declaration

I hereby declare that the contents and structure of this dissertation constitute my own original work and do not compromise in any way the rights of third parties, including rights related to the security of personal data.

Evelina Forno
2023

* This dissertation is presented in partial fulfillment of the requirements for a **Ph.D. degree** at the Graduate School of Politecnico di Torino (ScuDo).

Study and implementation of new computational paradigms exploiting neuromorphic hardware architectures

Evelina Forno

The increased use of machine learning (ML) techniques for predictive maintenance and real-time anomaly detection in embedded devices, IoT, and edge computing has led to challenges with limited power and memory. Neuromorphic systems offer low power consumption and effective data interchange, and spiking neural networks (SNNs) offer potential for offline and online learning. SNN-based solutions have shown comparable accuracy to ANNs while outperforming them in energy consumption.

This dissertation presents a general approach to generating neuromorphic models for IoT applicatios that can be used directly by users to facilitate the design process of solutions by exploring new computational paradigms. The structure of the thesis follows the flow of data in a neuromorphic pipeline, from raw sensory data to a completed application.

Embedded neuromorphic applications for edge deployment require interaction with sensors, including digital and event-based sensors. Event-based sensors offer extreme power efficiency and are likely to see widespread adoption for specialized tasks. However, digital sensors remain a more accessible and low-cost alternative for now, and their use is likely necessary for systems exploiting neuromorphic platforms to be deployed in the near future.

The author discusses encoding techniques for event-based and digital sensors, finding temporal contrast encoding to be a suitable first choice for most applications handling input data with an important time-varying component. The efficiency of coding techniques correlates with input frequency, so a technique must produce a sufficiently high spike count to stimulate all layers of the classification network. The work presents experiments involving a variety of SNN architectures, finding that for time-varying data in the spiking domain, recurrent neural networks like Lagrange Memory Unit and RSNN are more suitable due to their memory trace of past events. Model compression by edge pruning is also evaluated, obtaining significant reductions in synaptic activity without impacting performance.

Hyperparameter optimization is important for finding the optimal configuration of a network design. Results are reported for two HPO case studies and an exploration of the flexibility of the SpiNNaker system software to expand capabilities by implementing new placement and routing strategies. To evaluate the potential of neuromorphic hardware, the author performs a profiling of the parallel processing ability of the SpiNNaker board using a Message Passing Interface (MPI) implementation and the PageRank application as a benchmark. The performance of spiking neural networks on the neuromorphic Loihi chip is also compared to equivalent traditional and spiking classifiers on GPU, focusing on energy consumption per sample inference as a metric for assessing the advantages of a neuromorphic paradigm in this kind of application.

Finally, the author provides a summary of the research work by illustrating partial and complete neuromorphic pipelines, starting with a description of an interface for integrating the neuromorphic chip ODIN with a RISC-V coprocessor on a single SoC prototype. The processes used for neuromorphic classification of IoT time-varying signals are then compared through three case studies, showing the evolution of the pipeline from one project to the next, eventually encompassing the entire neuromorphic stack. The resulting method demonstrates a viable strategy for using neuromorphic tools in creating novel designs and implementations that offer outputs with limited error compared to those generated by applications implemented for general purpose architectures, while reducing execution time and power consumption, resulting in a net gain in adaptability for IoT and edge computing applications.